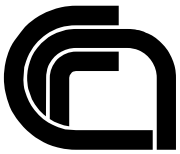

Consiglio Nazionale delle Ricerche

A criticism to Society (as seen by Twitter analytics)

S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, M. Tesconi

IIT TR-05/2014

Technical report

Marzo 2014

Istituto di Informatica e Telematica

A Criticism to Society
(as seen by Twitter analytics)

Stefano Cresci, Marinella Petrocchi,
Angelo Spognardi, Maurizio Tesconi
IIT-CNR, Pisa, Italy
E-mail: name.surname@iit.cnr.it

Roberto Di Pietro
Alcatel Lucent Bell Labs, France
IIT-CNR, Pisa, Italy
E-mail: roberto.di_pietro@alcatel-lucent.com

***Abstract*—Analytic tools are beginning to be largely employed, given their ability to rank, e.g., the visibility of social media users. Visibility that, in turns, can have a monetary value, since social media popular people usually either anticipate or establish trends that could impact the real world (at least, from a consumer point of view). The above rationale has fostered the flourishing of private companies providing statistical results for social media analysis. These results have been accepted, and largely diffused, by media without any apparent scrutiny, while Academia has moderately focused its attention on this phenomenon.**

In this paper, we provide evidence that analytic results provided by field-flagship companies are questionable (at least). In particular, we focus on Twitter and its "fake followers". We survey popular Twitter analytics that count the fake followers of some target account. We perform a series of experiments aimed at verifying the trustworthiness of their results. We compare the results of such tools with a machine-learning classifier whose methodology bases on scientific basis and on a sound sampling scheme. The findings of this work call for a serious re-thinking of the methodology currently used by companies providing analytic results, whose present deliveries seem to lack on any reliability.

I. Introduction

Social network sites (SNSs) keep playing a primary role in Internet users' life: they are used to access news, to keep in touch with friends, to share opinions and activities, to play collaborative games, and much more [1]. Many online tools investigate and aggregate the continuously produced information of SNS's users, and are also meant to increment user participation and engagement. Similarly, there is a flourishing availability of online tools as windows for data, news, goods to commercialize, primarily suited to attract and/or promote organizations and individuals: platforms to monitor user interactions (like *Facebook Insights, Woopra, PeerIndex, Klout*), integrated tools for simplifying and scheduling the communication with different SNSs (like *Buffer, HootSuite*), online software for customer relationship management (CRM) (like *Shoutlet, SproutSocial, Zoho*), just to cite a few. The typical online service for SNSs consists of an analytic engine that collects and process data from the target SNS. The output typically shows statistics on some parameter of the SNS (e.g., the number of subscribers, a monthly (or weekly)-based activity measurement of the network, etc.).

Nowadays, both traditional and online magazines and newspapers leverage such statistics to shed light on fashionable SNS's subjects (like VIPs or brands). Media often propose astonishing articles to their readers, aiming at starting gossip. Unfortunately, statistics could be provided without any provable guarantee of their reliability: they simply are supposed to work properly, either because of the reputation of the developer company or because media obtain remarkable revenues from them. Thus, very often, reliability of provided statistics is not considered as an issue, but just overlooked. One of the goals of this paper is focusing on the trustworthiness of the statistics offered by the SNSs analytic tools (from now on, "SNSs analytics"), with an eye to the inner models such tools rely on. As a running example, we concentrate on Twitter and we explore those analytics that are periodically in the spotlight for their impressive output: the count of Twitter *fake followers*.

Fake followers are Twitter accounts specifically created to inflate the number of followers of a target account, in order to increase its popularity and influence. As an example, during the 2012 US election campaign, bloggers and Twitter analysts noticed that the Twitter account of challenger Romney experienced a sudden jump in the number of followers, the great majority of them has been later claimed to be fake [2]. Since that date, the fake follower phenomena has grown more and more and the mainstream media have advertised a lot of analytics to spot Twitter fake followers ([3], [4], [5], [6]).

With a couple of clicks, the inquiring user can check the last report on her preferred VIP, to see how many of his followers are genuine, fake, or inactive. Problem is, in most cases, the scientific methodology behind the statistics reports is not available, the working principles are only sketched, and the results of the analysis is just accompanied by some discursive sentences listing high-level criteria that can help in discriminating between fake followers and genuine accounts. This approach is noticeably different from the scientific method of investigation, in which, upon the problem identification and the hypothesis for its solution, a series of observations, experiments, and relevant data are exploited to test such hypothesis, contextually providing other people with sound evidence and reproducible experiments.

Besides the risks of relying on an unpublished methodology, it is usually required to work via sampling, due to the huge number of involved elements (e.g., Barack Obama has more than 40 millions followers). However, this procedure can be error-prone if the sample is inadequate or biased. Having in mind the limitations that analytics may incur into, in this paper we investigate to which extent the Twitter analytics for detecting fake followers can be considered trustworthy. While

this issue has been already taken into consideration, e.g., [7], such attempts represent isolated cases and, to date, there is still a lack of scientific studies about the soundness of fake followers detection mechanisms.

The contributions of this work are as follows:

- we survey the most popular fake followers analytics, freely available on the Web: namely, the "Fakers" app of StatusPeople, the "Fake Follower Check" of Socialbakers, and "Twitteraudit". In particular, we report on their methodology and approach, from the point of the user of such tools;
- in order to analyse their scientific soundness, we go further in analysing such Fake Followers analytics, by performing a series of comparative experiments;
- finally, we compare the output of such tools with that of a machine-learning classifier whose methodology bases on scientific basis and on a sound sampling. Our experiments will reveal the lack of reliability of the analytics under investigation.

The remainder of this paper is as follows. Next section revises existing tools to spot Twitter fake followers and inactive accounts. Section III briefly presents the "Fake Project" fake follower classifier, designed and developed by the authors of this paper. In Section IV, we perform some experimental analysis, comparing the results of the considered tools. Finally, Section V concludes the paper.

II. Related Work

To the best of our knowledge, fake followers detection on Twitter has not been considered by the academic literature. Many interesting work focused on spam detection, see, e.g., [8], [9] or bot detection, see e.g., [10].

Instead, some social media companies have developed their own applications to spot fake followers. In this section, we detail three of the Twitter analytics that detect fake followers. The results of those analytics have been worldwide advertised by the Media, mainly highlighting the (supposed) high percentage of fake followers of VIPs, from politicians to singers. The three applications are: 1) the "Fakers" app by StatusPeople (referred by, for example, the Huffington Post [6], the "Fake Follower Check" by Socialbakers (mentioned, e.g., by the BBC [3] and the New York Daily News [4]) and "Twitteraudit" (as reported by, e.g., India Times [5]).

Apparently, the way these applications work is the same: on the website hosting the application, a Twitter user inputs the name of the Twitter account she wants to check. The application, then, requests the user to authorize itself to use her Twitter account and to access her profile, clearly listing the kind of operations it could do after that such authorization is granted. Finally, the application starts the analysis.

Considering the details provided by the official websites, the analysis process appears also common to all the three applications. Firstly, the application collects from the target account a list of names, that are followers of the target account. The number of collected names varies from application to application and it usually depends on the total number of followers of the target account. Secondly, the application randomly extracts from this list a sample of follower names. The size of the sample, again, varies from application to application. Then, the application collects and analyzes the information present on the Twitter profile of the sampled accounts. The way such a collection and analysis take place is the core of the application and its distinctive mark. Finally, a report of the analysis is output to the user, showing the percentage of fake followers that the application has detected for the target account.

A. Statuspeople.com

StatusPeople is a UK company, founded in August, 2011 and providing "a Social Media Management and Reporting Platform for business users". Their tools are made to ease the process of interaction of other companies with their customers, mainly focusing on social platforms like Facebook and Twitter. Fakers[1] is one of their tools that keeps being repeatedly mentioned by Media since its launch, dating July 2012. On the official web page[2] it is reported that, for an investigated account, "a sample of your follower data, up to 1,000 records depending on how 'popular' you are" is collected and assessed "against a number of simple spam criteria". In particular, the website notes that "on a very basic level spam accounts tend to have few or no followers and few or no tweets. But in contrast they tend to follow a lot of other accounts". Seeking for more details, we went through their blog. A post about the first update provides some hints: depending on the number of followers, the application collects up to 100K accounts and assess 1K of them. However, since 97% of Twitter accounts have less than 5K followers, the analysis of the application should consider a sound sample of the effective population. In a later post, dated 18 Oct 2012, after a modification to the Twitter API, there is a new update of the application: from 1K records across a follower base of up to 100K, the assessed records reduce to 700 across 35K. That one was the last post disclosing details about their application. However, the post[3] dating 26 Jul 2012 reports some concerns over the "Fakers" app, raised by bloggers and newspaper commentaries, pointing out that such low numbers would not accurately reflect accounts with a very large number of followers: for example, if an account with 100K genuine followers buys 10K fake followers, the application could show a 100% of fake, while the right percentage should be around 9% (namely 10K over a total of 110K). StatusPeople's spokesman answered that, for highly followed accounts, their application only provides insights about the "current follower activity", rather than a projection over the whole follower base. But, still there are no elements that clearly lists the features that a fake account should exhibit. During an interview, Rob Waller, the founder of StatusPeople, says that, among the several (unrevealed!)

[1] http://fakers.statuspeople.com/

[2] http://fakers.statuspeople.com/Fakers/FindOutMore/

[3] http://blog.statuspeople.com/Posts/Article/V/StatusPeople_Fakers_App_Shows_People_Care_About_Their_Follower_Quality/43

features they consider, the most meaningful one is the relationship between followers and friends of the account, i.e., "fake accounts tend to follow a lot of people but don't have many followers"[4].

Then, in November 2013, the StatusPeople blog presents the "Deep Dive" test, to provide more accurate scores for accounts with the highest number of followers (like, e.g., Katy Perry and Barack Obama): the Deep Dive samples the first 1.25 million records, assessing 33K followers of such a sample. The tool does not seem to be available to the users, but rather only accessible to StatusPeople internals. A blog post of January 2014, in reply to a Forbes article, reported different results with the Deep Dive tool, with respect to those obtained with Fakers, for accounts with a large number of followers: Barack Obama shifted from 70% fake to 45% fake, Lady Gaga from 71% to 39%, Shakira from 79% to 49%.

About the definition of an "active user" (that should be adopted in contrast with the notion of "inactive") Waller writes in a post that the active user is "someone who is engaging with the platform – producing and sharing content", and thus, s(he) is not simply someone who logs into the platform.

B. Socialbakers

Socialbakers is a social media analysis company, located in the Czech Republic and born in 2008. Its aim is to provide "solutions that allow brands to measure, compare, and contrast the success of their social media campaigns with competitive intelligence" . One of their free resources is the online tool Fake Follower Check (BETA)[5], launched in November 2012. The website provides details[6] about their methodology to detect fake (or "empty") followers. The rules have been updated more than one time over our period of observation. To date, when the following criteria are verified, than the account is considered fake:

- following/follower ratio = 50:1 (or more);
- more than 30% of the account's tweets use spam phrases (like *diet*, *make money*, *work from home*);
- the same tweets are repeated more than three times, even when posted to different accounts;
- more than 90% of the account's tweets are retweets;
- more than 90% of the account's tweets are links;
- the account has never tweeted;
- the account is more than two months old and still has a default profile image;
- the user did not fill in neither bio nor location and, at the same time, is following more than 100 accounts.

All the criteria have "a given number of points valuation" and if their combination exceeds "a certain number of points", the account is considered suspicious. Those accounts marked as suspicious are then tested against two more rules to verify if they are "inactive". One account is considered inactive by Socialbakers if:

- the account has posted less than 3 tweets;
- the last tweet is more than 90 days old.

Accounts that are neither suspicious, nor inactive, are considered genuine. The tool can be used ten times a day, considering "up to 2000 followers per account" and providing results with a declared "small error margin of roughly 10-15%".

No details are provided on how to weigh the satisfaction of each single criterion and the threshold above which an account is considered fake is not made explicit. Thus, we went through the official blog looking for hints about those numbers, but we were unable to find any further detail.

C. Twitteraudit.com

Twitteraudit.com hosts an application to check the percentage of fake followers of an account. According to the website, it is online since 2012 and managed by the Twitter users *@davc* and *@grossnasty*. Given each follower of an account, the application computes a score based on i) the number of its tweets, ii) the date of the last tweet, and iii) the ratio of followers to friends, "taking a random sample of 5K Twitter followers". There are no details on how the score is computed. In addition to the percentage of the resulting fake followers, the audit also outputs three charts: the first chart describes how Twitteraudit considers the checked account (fake, not sure, real), the second one reports the "quality score" per follower (with no explanation on what a "quality score" is) and the third chart details the "real points" per follower, with a maximum scale of 5. According to this output, we can argue that the three criteria used to evaluate the score can sum up to five.

Twitteraudit.com makes also available a Fake Followers Chrome extension that shows the fake follower percentage when visiting the account page of a Twitter user.

D. Observations and remarks

As we have seen, the analytics under investigation do a small sampling of the follower list of the target account and do not clearly explain how they combine the adopted criteria for the fake identification. What is more, the sampling is not drawn uniformly at random, but it seems to be time-dependent. That is, the followers taken into consideration are just the latest ones to have joined—where "latest" differs from engine to engine, but it is always a fixed number, unrelated to the total number of followers. Here we want to stress how those two elements can affect the results of the assessment.

Firstly, we recall some simple statistical notions for estimating the proportion p of a given population that holds a given property [11]. The estimator $\hat{p} = X/n$, where X is the number of samples positive to the property and n is the size of the sample. Assuming i) a non biased sample, ii) an independent choice of the samples and iii) a perfect test of the property, the estimator can be approximated by a normal distribution. We can evaluate its standard error σ with the variance $\sigma^2 = \frac{\hat{p}*(1-\hat{p})}{n}$. Then, the confidence interval is $\hat{p} \pm Z_\alpha \sigma$, where Z_α is the critical value that depends on the confidence level: with a confidence level of 0.95 $Z_\alpha = 1.96$, while for 0.99 $Z_\alpha = 2.58$.

[4]http://neonfresh.com/status-people/

[5]http://www.socialbakers.com/twitter/fakefollowercheck

[6]http://www.socialbakers.com/twitter/fakefollowercheck/methodology

However, when the number of followers is larger than 10K, our assessed analytics seem to have all the three assumptions flawed. Firstly, the assumption i) is not satisfied, since the sample is not unbiased: all the applications get the sample not from the whole list of followers, but only from the ones that started following the account more recently (this will be shown in Section IV). Secondly, the assumption ii) not fulfilled either, since the choice of the samples is not independent among the followers: the applications compose the sample picking from n followers (n varies between the applications), instead of N, that is the whole population. Such issues greatly impact on the representativeness of the sample. And finally, regarding assumption iii), the three companies provide no evidences about the correctness of their fake detector engine.

III. The Fake Project engine

In a recent work [12], we have proposed a statistically sound fake detection engine, with an open source classifier. By contrast to the surveyed applications, our engine uses the whole list of followers to perform the sampling and clearly shows how to reproduce the detection engine. It considers as "inactive" any follower that has never tweeted or whose last tweet is older than 90 days. It considers "fake" followers those accounts satisfying a series of properties, not reported here for the sake of brevity, but fully presented in [12].

The fake classifier[7] (*FC* from here on) has been designed within the CNR's "Fake Project"[8], upon testing known methodologies for bot and spam detection on a gold standard of Twitter accounts, where fake followers, inactive, and genuine accounts were *a priori* known. In particular, we have applied to the accounts in our reference set algorithms based on 1) single classification rules proposed by [13], [14], [15], and 2) feature sets proposed in the literature by [8], [9]. The outcome of the analysis led us to conclude that fake followers detection deserves specialized mechanisms: in particular, algorithms based on classification rules do not succeed in detecting the fakes in our reference dataset, while better results were achieved by relying on those features proposed by Academia for spam accounts detection. Based on the features and rules that best behave in detecting fake followers, we went further by looking for an "optimized" classifier that considers also the evaluation costs of such features and rules. Thus, we have quantified their crawling cost and we built a set of optimized classifiers that make use of the more efficient features and rules, in terms both of crawling cost and fake followers detection capability. The interested reader can find all the details in [12].

IV. Experimental Analysis

In this section, we report on two kind of experiments we have performed. The first experiment aims at testing the Twitter API used to gather data of Twitter accounts: our hypothesis is that the API that requests the list of followers of an account (hereafter, the target) reports the followers in the reverse order with respect to "following time". This would mean that the list of the first 1000 followers returned by Twitter is actually the list of the last 1000 accounts that started following the target.

The second experiment directly involves the Twitter analytics surveyed in Section II. We run them over a set of Twitter accounts (with a different number of followers) and collect statistics on the execution of the analysis, i.e., the results of the analysis and how much time has been needed to produce them. Based on the outcome of the experiment, we then discuss the reliability and soundness of the analytics under investigation.

API type	*elem.*×*request*	*max requests*×*min.*
GET followers/ids	5000	1
GET friends/ids	5000	1
GET users/lookup	100	12
GET statuses/user_timeline	200	12

Table I: Twitter APIs: type and limitations to API calls. The complete list at: https://dev.twitter.com/docs/rate-limiting/1.1

A. The testbed

We select Twitter accounts with a *low* (10K or less), *average* (>20K and <100K), and *high* (>100K) number of followers. In particular, for the *average* class of accounts, our aim is to evaluate the response time of the analytics. For those classes, we then choose accounts that would have unlikely been cached for fake follower detection (i.e., accounts with a relevant number of followers, but not belonging to worldwide popular characters). Thus, for the *average* class, we identify the Twitter accounts of thirteen individuals quite popular in Italy, with an average number of followers of about 50K.

For the *low* class, we select the accounts of the analytics developers, namely @*davc* and @*grossnasty* of Twitteraudit (TA), @*RobDWaller* of StatusPeople (SP) and @*janrezab12*, CEO of Socialbakers (SB), that have 10K of followers or less. Finally, we select the accounts of three well-known politicians, i.e., Obama, Cameron, and Hollande, as representatives of the *high* class of accounts.

B. Twitter API analysis

Twitter has released a number of APIs that can be used by developers to implement a set of various applications[9]. These applications, with a set of permissions granted and the use of the available APIs, automatically and quickly interact with a target. In particular, the minimum set of APIs to run a fake follower check on a target are *GET followers/ids* and *GET users/lookup*. *GET followers/ids* fetches the list of followers of the target while *GET users/lookup* gives access to the account information of its followers. Other APIs are provided to acquire more details about the target accounts, such as the timelines (restricted however to the last 3200 tweets of an account). To protect Twitter from abuse, the number of API calls allowed per minute is limited. In Table I, we report the

[7] http://wafi.iit.cnr.it/fake/fake/app

[8] http://wafi.iit.cnr.it/theFakeProject/

[9] http://dev.twitter.com

		seconds			
Twitter profile	*followers*	*FC*	*TA*	*SP*	*SB*
@giovanniallevi	13900	187	55	27	12
@StefanoBollani	22300	188	52	22	11
@Federugby	30300	193	40	31	13
@Zerolandia	33500	193	51	32	9
@pinucciotwit	35500	192	3	2	13
@mvbrambilla	36900	188	45	2	8
@PChiambretti	40500	198	45	23	9
@pierofassino	61500	203	52	3	10
@Lbarriales	69900	212	50	27	7
@PC_Chiambretti	70900	214	43	31	9
@herbertballeri	72300	217	54	24	10
@Flaviaventosole	75400	210	49	27	9
@RudyZerbi	79700	216	49	26	10

Table II: Response time to first analysis request

maximum number of calls allowed per minute, which directly impacts on the time needed to complete the data acquisition process. This implies that collecting data of accounts with a very large numbers of followers can be extremely time consuming. For example, for our tests we gathered data from the whole set of followers of President Obama. This required a total time of around 27 days.

Our analysis is mainly based on *GET followers/ids* API, used to fetch the target's list of followers. We aim at verifying that this API returns the followers of the target, starting from the ones that became followers more recently. If so, those analytics only fetching a subset of followers using the above API, would actually consider only the newest followers, leading to the analysis of a biased sample of followers (see Section II-D).

To verify our thesis, for every account of our *average* testbed, we saved the whole list of followers, together with their position in the list, once per day. We then compared the lists day by day, observing the positions in the lists of the new users that started following our target accounts. We verified that all the new entries in all the lists of followers were always added at the end. This confirmed our thesis.

C. Fake Follower analysis response time

The response time to the first analysis request of the thirteen Italian accounts under investigation is as in Table II (we repeated the analysis several times). From the results, it appears clear that some of the analytics have some results already computed. The analysis response time of the Fake Project engine (*FC*, *cfr* Section III) is always greater than 180 seconds: this is because, first, it requests the complete list of followers and, then, it requests the profiles of the sampled ones. Moreover, to be statistically sound, the sample size is always 9604, to guarantee a confidence level of 95%, with a confidence interval of 1%. Such response time is consistent with the time needed to invoke the Twitter API for the required number of times, given the limitations presented in Table I. Twitteraudit (*TA*) is the only that explicitly reports the assessment date: for example, the results of *@pinucciotwit* were output after only 3 seconds since it was evaluated "7 months ago". On the contrary, observing the response time of StatusPeople (*SP*), it is evident when the results were cached: while the average time was around 25 seconds, the reports of three accounts (i.e., *@pierofassino, @mvbrambilla, @pinucciotwit*, two politicians and one opinionist, respectively) were displayed after 2 seconds only (without mentioning if the analysis had been performed in advance). Socialbakers (*SB*), instead, does not seem to have performed any caching of the results, since all the analysis took almost the same response time. It is worth noticing that, for the subsequent requests of analysis on the same accounts, all the tools output the results in less than 5 seconds. Again, only *TA* reported the last assessment time, while *SB* and *SP* do not explicitly say if the analysis is real-time, or if the results have been cached.

D. Fake Follower analysis results

For the four applications under investigation (the three Twitter analytics and the Fake Project engine *FC*), Table III reports the number of inactive, fake, and genuine followers of the targets. Overall, we may observe that there is a general disagreement on such results. Excluding Twitteraudit, that does not consider inactive followers, the Socialbakers Fake Follower Check tends to recognize less inactive followers than the *FC* classifier and StatusPeople Fakers. However, there are examples for which the disagreement is high also between *FC* and *SP*: for *@mvbrambilla, @PC_Chiambretti* and *@RudyZerbi* the inactive followers are, respectively, 75.7, 97 and 83.8 for *FC* and 42, 48 and 44 for StatusPeople. Twitteraudit and Socialbakers results are similar for number of genuine followers. Lastly, it seems that the more followers a target has, the less the fake followers analytics agree.

We found an interesting issue regarding the following two accounts: *@PC_Chiambretti* and *@PChiambretti*. Apparently, both belong to Piero Chiambretti, a well-known Italian comedian and TV presenter. The first account was created on December 2011 to advertise an upcoming TV show and has tweeted only 13 times. On the contrary, the second account was created on October 2012 and has tweeted 2500 times: since its launch, the account continuously tweets personal opinions of the comedian and posts his pictures. *FC* recognized that, on a sample of 9604 (taken from the whole 70900) *@PC_Chiambretti*'s followers, 97% of them (9314) are inactive. The analytics under investigation, instead, reported much lower percentages: Twitteraudit only recognized 55% of fakes, StatusPeople 48% inactive accounts and 44% fakes, while Socialbakers 17% inactive accounts and 35% fakes. Such values seem to confirm that the choice to select the sample considering only the latest followers can lead to inaccurate results.

Looking at the last three accounts in Table III, which have the highest number of followers, we can see that the differences in results further increase. Twitteraudit, Socialbakers, and *FC* assess the highest percentage of genuine followers for Cameron, contrasting with StatusPeople, which only finds 35% of genuine followers. Overall, *SP* Fakers minimizes

		Fake Classifier			*Twitteraudit**		*Statuspeople*			*Socialbakers*		
Twitter profile	*followers*	*inact.*	*fake*	*good*	*fake*	*good*	*inact.*	*fake*	*good*	*inact.*	*fake*	*good*
@RobDWaller	929	25	1.4	73.6	7	93	28	0	72	0	0	100
@davc	2971	13.5	4.1	82.4	14	86	26	3	71	0	4	96
@grossnasty	3344	12.9	4	83.1	4	96	26	3	71	0	2	98
@janrezab	10800	18.4	2.2	79.4	11	89	27	3	70	2	2	96
@giovanniallevi	13900	44.3	9.9	45.8	34	66	58	18	24	5	27	68
@StefanoBollani	22300	27.8	12.8	59.4	29	71	49	11	40	12	11	77
@Federugby	30300	46.5	15.5	38	42	58	51	33	16	9	33	58
@Zerolandia	33500	69.2	7.3	23.5	63	37	55	35	10	24	25	51
@pinucciotwit	35500	30	6.3	63.7	28	72	25	13	62	7	15	78
@mvbrambilla	36900	75.7	6.5	17.8	47	53	42	30	28	9	34	57
@PChiambretti	40500	31.6	21.7	46.7	36	64	56	22	22	13	19	68
@pierofassino	61500	77.9	4.6	17.5	46	54	39	39	22	14	31	55
@Lbarriales	69900	49.5	20.6	29.9	48	52	57	32	11	13	21	66
@PC_Chiambretti	70900	97	1.2	1.8	55	45	48	44	8	17	35	48
@herbertballeri	72300	46	10.4	43.6	48	52	56	22	22	14	20	66
@Flaviaventosole	75400	46.4	12.8	40.8	39	61	46	33	21	12	29	59
@RudyZerbi	79700	83.8	5.9	10.3	35	65	44	33	23	8	26	66
@David_Cameron	595K	24	11.7	64.3	19.5	80.5	17	48	35	10	14	76
@fhollande	608K	63.6	5.3	31.1	64.3	35.7	35	44	21	44	14	42
@BarackObama	41M	57.1	8.5	34.4	51.2	48.8	40	41	19	43	12	45

Table III: Fake follower analysis results (*twitteraudit does not consider inactive followers)

the number of genuine followers, compared to the number reported by the other tools. *SB* and *SP* show substantially lower percentages of inactive accounts than *FC*, over all the three politicians. As in the case of *@PC_Chiambretti*, this difference can be explained considering that *SB* and *SP* only analyze a small set of the newest followers and new followers are less likely to be inactive than long-term followers. Therefore, a sampling among the newest followers is likely to show lower scores for inactive accounts than the scores computed from a sample over all the followers. Overall results for Obama, Cameron and Hollande seem to further confirm the doubts about the low reliability of the evaluated closed-source tools.

We can conclude two main facts: firstly, we can confirm that the choice to use a small sample of the last followers can lead to very inaccurate results. Secondly, the general disagreement confirms that the use of a closed-source fake detector engine, that acts like a "black box", produces results that cannot be considered trustworthy. The adoption of criteria that are clearly stated and open to discussion should be eventually preferred. The Fake Project *FC* engine is based on publicly available criteria and the training dataset is available on request.

V. Conclusions

In this paper, we surveyed and analyzed three quite popular Twitter analytics, counting the number of fake followers of target accounts. We proved that these tools do not properly fulfill the basic assumptions for an unbiased sampling. Also, our experiments showed how the results of the tools, run over the same set of target accounts, are generally misaligned, leading to the suspect of scarce reliability. Such insights, other than being interesting of their own, pave the way for further investigation. Towards this direction, we propose the use of more reliable tools, making use of proper statistical sampling and with a disclosed methodology of analysis.

References

[1] D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[2] NBC News (online ed.), "Romney Twitter account gets upsurge in fake followers, but from where?" in *http://goo.gl/WC4Ucu*, Aug. 2012, last checked Dec. 27, 2013.

[3] BBC - Newsbeat, "Katy Perry overtakes Justin Bieber on Twitter followers," in *http://goo.gl/SB0Yyq*, Nov. 2013, last checked Feb. 19, 2014.

[4] Zayda Rivera - New York Daily News, "Justin Bieber may not be 'The king of Twitter' anymore! Report shows half his followers are fake," in *http://goo.gl/f8Eabh*, Apr. 2013, last checked Feb. 19, 2014.

[5] The Economic Times - India Times, "Twitter Audit separating fake followers from real followers," in *http://goo.gl/8kFNaK*, Sep. 2013, last checked Feb. 19, 2014.

[6] Grant Stern - The Huffington Post, "Can Twitter Curb Growing Black Market for Fake Followers After IPO?" in *http://goo.gl/XkocvE*, July 2013, last checked Feb. 19, 2014.

[7] Jim Dougherty - leaderswest.com, "How accurate are fake Twitter follower tools?" in *http://goo.gl/09u3v4*, May. 2013, last checked Feb. 19, 2014.

[8] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *26th Annual Computer Security Applications Conference*, ser. ACSAC '10. ACM, 2010, pp. 1–9.

[9] C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 8, pp. 1280–1293, 2013.

[10] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" *IEEE Trans. Dependable Sec. Comput.*, vol. 9, no. 6, pp. 811–824, 2012.

[11] W. A. Fuller, *Sampling Statistics*. Wiley, Sep. 2009.

[12] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "A Fake Follower Story: improving fake accounts detection on Twitter," IIT-CNR, Tech. Rep. TR-03, 2014, submitted.

[13] M. Camisani-Calzolari, "Analysis of Twitter followers of the US Presidential Election candidates: Barack Obama and Mitt Romney." in *http://digitalevaluations.com/*, Aug. 2012.

[14] SocialBakers, "Fake follower check," in *http://goo.gl/chWn0*, last checked Dec. 27, 2013.

[15] Stateofsearch.com, "How to recognize Twitterbots: 7 signals to look out for." in *http://goo.gl/YZbVf*, Sep. 2012, last checked Dec. 27, 2013.